The Discovery of Initial Fluxes of Metabolic P Systems

Roberto Pagliarini, Vincenzo Manca

Verona University
Computer Science Department
Strada Le Grazie 15, 37134 Verona, Italy.
{roberto.pagliarini,vincenzo.manca}@univr.it

Summary. A central issue in systems biology is the study of efficient methods to infer fluxes of biological reactions starting from experimental data. Among the different techniques proposed in the last years, in the theory of *Metabolic P systems Log-Gain* principles have been introduced, which prove to be helpful for deducing biological fluxes from temporal series of observed dynamics. However, crucial tasks remain to be performed for a complete suitable application of these principles. In particular the algebraic systems introduced by the Log-Gain principles require the knowledge of the initial fluxes associated with a set of biochemical reactions. In this paper we propose an algorithm for estimating initial fluxes, which is tested in two case studies.

1 Introduction

In the last years, the problem of reverse-engineering of biological phenomena from experimental data has spurred increasing interest in the scientific communities. For these reasons, many computational models inspired from biology have been given. Among these models, the *Metabolic P systems* [9, 10], shortly *MP systems*, proved to be relevant in the analysis of dynamics of biochemical processes [4, 12, 14, 13]. MP systems intend to model metabolic systems, that is, structures where matter of different types is subject to reactions, or transformations of various types. The importance of these computational models is their potential applicability to the reverse-engineering problem of biological phenomena. In fact, the MP systems introduce a theory, called *Log-Gain* [8], intrinsically related to the structure of these computational models.

This theory provides a method for constructing MP models of real phenomena from time-series of observed dynamics. In fact, given a real system which can be observed in its evolution, then almost all the elements occurring in the definition of MP system can be, in principle, deduced by macroscopic observations of the system [9].

The only component which cannot be directly measured is a set of regulation functions which state the reaction fluxes, that is, the amount of reactants transformed by the reactions at any state of the system. These functions depend on internal microscopic processes on which molecules are involved. However, Log-Gain theory provides a way for deducing them from time-series of the states of a given system along a sufficient number of observation instants.

A key point for achieving this task consists in the discovery of the fluxes associated to the passage of a metabolic system from two initial observation instants. In this paper an algorithm is proposed for estimating these initial fluxes from few steps of observation.

The present paper is organized as follows. Section 2 is devoted to the definition of Metabolic P Systems. Section 3 briefly recalls the Log-Gain theory. In Section 4 we describe the algorithm that solves our problem. Section 5 reports some experimental results obtained by the new framework. Some further remarks and directions for future researches are discussed in the last section.

2 Metabolic P Systems

MP systems are a special class of dynamical systems (the reader can find some details concerning dynamical aspects of MP systems in [11]), based on *P systems* [3, 16, 17, 18], which are related to metabolic processes. MP systems are essentially constituted by multiset grammars where rules are regulated by specific functions depending on the state of the system. From a Membrane Computing point of view, MP systems can be seen as deterministic mono-membrane P systems where the transitions between states are calculated by a suitable recurrent operation. In an MP system the overall variation, in a macroscopic time interval, of the whole system under investigation is considered. In this manner, the evolution law of the system consists in the knowledge of the contribution of each reaction in the passage between any two instants separated by such an interval. Therefore, dynamics is given at discrete steps, and at each step, it is ruled by a partition of matter among the reactions transforming it. The principle underlying the partitioning is called *mass partition principle*. This principle replaces the *mass action law*[1] of ODE systems. The mass partition principle defines the transformation rate of object populations rather than single objects, according to a suitable generalization of chemical laws [9].

[1] The foundation of this law is the theory of molecular collisions. The first formulation of this law, formulated by Waage and Guldberg [21], is the following: *"the rate of any given chemical reaction is proportional to the product of the concentrations of the reactants"*.

2.1 MP systems: a formal definition

The following definition introduces the MP systems in a formal way ($\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ respectively denote the sets of natural, integer, and real numbers).

Definition 1 (MP system) *An MP system M is specified by the following construct:*

$$M = (X, R, V, H, \Phi, \nu, \mu, \tau, \delta)$$

where X, R and V are finite disjoint sets, and moreover the following conditions hold, with $n, m, k \in \mathbb{N}$:

- *$X = \{x_1, x_2, \ldots, x_n\}$ is a finite set of substances. This set represents the types of molecules;*
- *$R = \{r_1, r_2, \ldots, r_m\}$ is a finite set of reactions. A reaction r is a pair of type $\alpha_r \rightarrow \beta_r$, where α_r identifies the multiset of the reactants (substrates) of r and β_r identifies the multiset of the products of r (λ represents the empty multiset). The stoichiometric matrix $\mathbb{A}$ of a set R of reactions over a set X of substances is $\mathbb{A} = (\mathbb{A}_{x,r} \mid x \in X, r \in R)$ with $\mathbb{A}_{x,r} = |\beta_r|_x - |\alpha_r|_x$, where $|\alpha_r|_x$ and $|\beta_r|_x$ respectively denote the number of occurrences of x in α_r and β_r. Of course, a reaction r can be seen as the vector $r = (\mathbb{A}_{x,r} \,|x \in X)$ of $\mathbb{Z}^n$. We also set $R_\alpha(x) = \{r \in R \mid x \in \alpha_r\}$, $R_\beta(x) = \{r \in R \mid x \in \beta_r\}$, and $R(x) = R_\alpha(x) \cup R_\beta(x)$;*
- *$V = \{v_1, v_2, \ldots, v_k\}$ is a finite set of parameters (such as pressure, temperature,...);*
- *$H = \{h_v \mid v \in V\}$ is a set of parameters evolution functions. The function $h_v : \mathbb{N} \rightarrow \mathbb{R}$ states the value of parameter v, and $H[i] = (h_v(i) \,|v \in V)$;*
- *$\Phi = \{\varphi_r \mid r \in R\}$ is the set of flux regulation maps, where, for each $r \in R$, $\varphi_r : \mathbb{R}^{n+k} \rightarrow \mathbb{R}$. Let $q \in \mathbb{R}^n$ be the vector of substances values and $s \in \mathbb{R}^k$ be the vector of parameters values. Then $(q, s) \in \mathbb{R}^{n+k}$ is the state of the system. We set by $U(q, s) = (\varphi_r(q, s) \mid r \in R)$ the flux vector in the state (q, s);*
- *ν is a natural number which specifies the number of molecules of a (conventional) mole of M;*
- *μ is a function which assigns, to each $x \in X$, the mass $\mu(x)$ of a mole of x (with respect with to some measure units);*
- *τ is the temporal interval between two consecutive observation steps;*
- *$X[i] = (x_1[i], x_2[i], \ldots, x_n[i])$, for each $i \in \mathbb{N}$, is the vector of substances values at the step i, and $X[0]$ are the initial values of substances. The dynamics $\delta : \mathbb{N} \rightarrow \mathbb{R}^n$ of the system is completely identified by the following recurrent equation, called Equational Metabolic Algorithm shortly EMA:*

$$X[i+1] = \mathbb{A} \times U(X[i], H[i]) + X[i] \tag{1}$$

where $\mathbb{A}$ is the stoichiometric matrix of reactions having dimension $n \times m$, while $\times$, $+$, are the usual matrix product and vector sum. We denote by $EMA[i]$ the system (1) at the step i. By using the formulation introduced above it is simple to note that we can obtain the vector $X[i+1]$ from vectors $X[i]$ and $U(X[i], X[i])$.

3 Log-Gain Theory: A Brief Recall

The starting point of the Log-Gain theory for MP systems [20] is the *Allometry Law* [2, 6] which assumes a proportion between the relative variations of the fluxes of a reaction r and the sum of relative variations of *tuners* of r, that is, magnitude influencing r.

The relative variation of a variable x is given, in differential notation and with respect to the time variable t, by $d(\lg x)/dt$. This explains the term "Log-Gain".

Given a dynamics of an MP system, we will use the following simplified notations, for $i \in \mathbb{N}$, and $r \in R$:

$$u_r[i] = \varphi_r(X[i], H[i]) \quad \text{and} \quad U[i] = (u_r[i] | r \in R) \tag{2}$$

Assuming to know the vectors $X[i]$ and $X[i+1]$, the equation (1) can be rewritten in the following form, which we called $ADA[i]$ (Avogadro and Dalton Aggregation [10]):

$$X[i+1] = \mathbb{A} \times U[i] + X[i] \tag{3}$$

The formula (3) expresses a system of n equations and m variables (n is the number of substances and m the number of reactions) which is assumed to have maximal rank. This supposition is not restrictive. In fact, if it does not hold the rows which are linearly depend on other rows are removed. Formally $ADA[i]$ is the same to system $EMA[i]$ introduced in Section 2. However, these two systems have dual interpretations. In fact, in $EMA[i]$, the vectors $U[i]$ and $X[i]$ are known, and the vector $X[i+1]$ is computed by means of them, while in $ADA[i]$, the vector $X[i+1] - X[i]$ is known and $U[i]$ is computed by solving the system, as we will see by formula (6).

Usually, in a biochemical phenomenon, the number of reactions is greater than the number of substances, and this means that the system (3) has more than one solution. Therefore, fluxes cannot be univocally deduced by means of ADA. The following principle [8] allows us to add more equations to the above system in order to get a univocally solvable system which could provide the flux vector.

Definition 2 (Discrete Log-Gain) *Let $(z[i] \,|\, i \in \mathbb{N})$ a real valued sequence. Then, the discrete log-gain of z is given by the following equation:*

$$Lg(z[i]) = \frac{z[i+1] - z[i]}{z[i]} \tag{4}$$

Principle 1 (Log-Gain regulation) *Let $U[i]$, for $i \geq 0$, be the vector of fluxes at step i. Then the Log-Gain regulation can be expressed in terms of matrix and vector operations:*

$$(U[i+1] - U[i])/U[i] = \mathbb{B} \times L[i] \quad + C \otimes P[i+1] \tag{5}$$

where:

- $\mathbb{B} = (p_{r,z} \,|r \in R, z \in X \cup V\,)$ *where* $p_{r,z} \in \{0,1\}$ *with* $p_{r,z} = 1$ *if* z *is a tuner of* r *and* $p_{r,z} = 0$ *otherwise;*
- $L[i] = (Lg(z[i]) \,|z \in X \cup V\,)$ *is the column vector of substances and parameters log-gains ;*
- $P[i+1]$ *is a column vector of values associated with the reactions and called (Log-Gain) offsets at step* $i+1$*;*
- $C = (c_r \,|r \in R)$*, where* $c_r = 1$ *if* $r \in R_0$*, while* $c_r = 0$ *otherwise, and* R_0 *is a subset of reactions having the Covering Offset Log Gain Property, that is, it is a set of* n *linear independent vectors of* $\mathbb{Z}^n$*;*
- $\times$ *denotes the usual matrix product;*
- $+, -, /, \otimes$ *denote the component-wise sum, subtraction, division and product*[2]*.*

If we assume to know the flux unit vector at step i and put together the equations (5) and (3) at steps i and $i+1$ respectively, we get the following linear system called *Offset Log-Gain Adjustment* module at step i, shortly $OLGA[i]$, in which the number of variables (here reported in bold font) is equal to the number of equations:

$$\begin{aligned} \mathbb{A} \times \mathbf{U[i+1]} &= X[i+2] - X[i+1] \\ (\mathbf{U[i+1]} - U[i])/U[i] &= \mathbb{B} \times L[i] \quad + \quad C \otimes \mathbf{P[i+1]} \end{aligned} \tag{6}$$

Now, if the vectors $X[i]$ and $V[i]$, for $0 \leq i \leq l$, where $l \in \mathbb{N}$, are obtained by experimental measures, then it is possible to solve $OLGA[i]$ for $i = 0, \ldots, l-1$, obtaining the vector $U[i]$ for $i \in [1, l-1]$.

4 An Algorithm for the Estimation of Initial Metabolic Fluxes

The method described in the previous section assumes the knowledge of the initial values of fluxes.

Problem 1 (Initial Fluxes Problem) *Given* $X[0]$ *and* $H[0]$*, find a flux vector* $U[0]$ *such that it satisfies the initial dynamics, that is:*

$$X[1] \cong \mathbb{A} \times U[0] + X[0]$$

where $\cong$ *means approximate equality.*

The algorithm given below circumvents the Initial Fluxes Problem by using the knowledge about the dynamics in the first evolution steps in order to approximate the amount of substances which is not transformed, we call *inertia* of the system (at a given step).

[2] Given two $n \times m$ matrices A and B, the operation $A \otimes B$ involves the action of multiplying component-wise each element of A by the corresponding element of B.

4.1 The proposed algorithm

Our approach is based on the assumption that if the inertia of each substance is known, then only a part of substances has to be partitioned among the reactions which require to consume them. The main steps of the algorithm are described in the following of this section.

Step 1.

The goal of the first step is to evaluate grossly the initial fluxes at the step 0 by assuming that they are proportional to the reactants, that is, for all $r \in R$:

$$\hat{u}_r[i] = k_r y_r[i] \tag{7}$$

where $k_r \in \mathbb{R}$, and $y_r[i]$ is the product of all substance quantities, at the step i, which are reactants for r. We suppose that if $\alpha_r = \lambda$ then $y_r = 1$, and we set

$$\hat{U}[i] = (\hat{u}_r[i] \mid r \in R) \tag{8}$$

For example, in a metabolic system with three kinds of substances, a, b, c, and with a set of reactions given in the first column of the Table 1, the relationships among the fluxes of these reactions and their reactants are reported in the second column of the Table 1.

Let us consider the following system, called *Local-Stoichiometric Module* at the

Reactions	Maps
$r_1 : a \to bc$	$k_{r_1} a$
$r_2 : b \to a$	$k_{r_2} b$
$r_3 : c \to ab$	$k_{r_3} c$
$r_4 : c \to cc$	$k_{r_4} c$

Table 1. Reactions and their flux regulation maps of the Local-Stoichiometric Module.

step i:

$$x[i+1] - x[i] = \sum_{r \in R(x)} \mathbb{A}_{x,r} \hat{u}_r[i] \quad \forall x \in X \tag{9}$$

If we assume that the constants k_r, with $r \in R$, do not sensibly change in few steps, then by applying the system (9) for a sufficient number of steps we can obtain a square linear system of dimension m having maximum rank. In the example reported in Table 1, we have a Local-Stoichiometric Module of 3 equations which initially has 4 unknowns. It has rank 3. At the second iteration of this module we get other 3 equations and the rank of Local-Stoichiometric Module is maximum. Thus, we can obtain a system of equation having unique solution. In general, if we start with the Local-Stoichiometric Module at the step 0 then we can compute the vector $\hat{U}[0] = (\hat{u}_r[0] \mid r \in R)$ by applying the Local-Stoichiometric module a suitable number of steps.

Step 2.

The aim of this step is to approximate the inertia of the system. We split this step in two sub-steps. In the first one we take n linear independent reactions, obtaining a set R_0, according to the Covering Offset Log Gain Property. Then, we use the set R_0 to obtain an $OLGA[1]$ module, with $U[0] = \hat{U}[0]$, where $\hat{U}[0]$ is the vector of fluxes computed in the previous step. We will indicate with $U^*[1] = (u_r^*[1] \mid r \in R)$ the solution of this system. However, if some elements of this vector is a negative real value, then we choose others n linear independent reactions and reapply the procedure above describe (it easy to prove that a positive vector must exist).

In the second sub-step we compute, for each $x \in X$, the inertia, indicated by $\bar{x}$, by applying the following equation:

$$\bar{x}[1] = x[1] - \sum_{r \in R_\alpha(x)} u_r^*[1], \quad \forall x \in X \tag{10}$$

Step 3

In the last step we obtain the vector of fluxes at the evolution step 1 by solving an optimization problem. In fact, the vector $U^\circ = (u_r^\circ \mid r \in R)$ we search has to be a strictly positive vector of $\mathbb{R}^m$ (positive in each component) which satisfies the following n equations:

$$x[1] - \bar{x}[1] = \sum_{r \in R_\alpha(x)} u_r^\circ[1], \quad \forall x \in X \tag{11}$$

and it is bounded, for each component, by the following constraint:

$$u_r^\circ[1] \leq \begin{cases} \min\left\{ \frac{x_j[1] - \bar{x}_j[1]}{|\alpha_r|_{x_j}} \mid x_j \in \alpha_r \right\} & \text{if} \quad \alpha_r \neq \lambda \\ k_r & \text{if} \quad \alpha_r = \lambda \end{cases} \tag{12}$$

and such that

$$U^\circ = \min_{\xi \in \mathbb{R}^m} \|\mathbb{A} \times \xi - (X[2] - X[1])\| \tag{13}$$

5 Experiments

In this section, in order to evaluate the performance of our algorithm, we apply it to two case studies: *i)* a synthetic oscillatory metabolic system, and *ii)* the Belousov-Zhabotinsky reaction [1, 7, 19, 22].

Reactions	Flux regulation maps
$r_1 : a \to aa$	$\varphi_1 = k_1 a/(k_1 + k_2 c + k_4 b + k_6)$
$r_2 : a \to b$	$\varphi_2 = k_2 ac/(k_1 + k_2 c + k_4 b + k_6)$
$r_3 : b \to \lambda$	$\varphi_3 = k_3 b/(k_3 + k_6)$
$r_4 : a \to c$	$\varphi_4 = k_4 ab/(k_1 + k_2 c + k_4 b + k_6)$
$r_5 : c \to \lambda$	$\varphi_5 = k_5 c/(k_5 + k_6)$
$X[0] = (100 \quad 100 \quad 1)$	$k_1 = k_3 = k_5 = 4, k_2 = k_4 = 0.02, k_6 = 100$

Table 2. Sirius' reactions and maps.

5.1 A synthetic metabolic system

Let us consider the synthetic non-cooperative metabolic system called Sirius and given in Table 2 [9]. Firstly, we generate the dynamics of this model for 1000 steps by using the flux regulation maps of Sirius. Then, we use our algorithm to approximate the vector of fluxes $U^{\circ}[1]$ at the evolution step 1. Starting from $U^{\circ}[1]$, by applying $OLGA[i]$ for $i = 1, 2, \ldots, 900$, we deduce the vectors $U[i]$, for $i = 2, 3, \ldots, 899$, according to the Log-Gain theory. Figure 1 shows the fluxes relative to the dynamics of Sirius initially generated, while the Figure 2 shows the inferred fluxes. These results show an almost complete accordance.

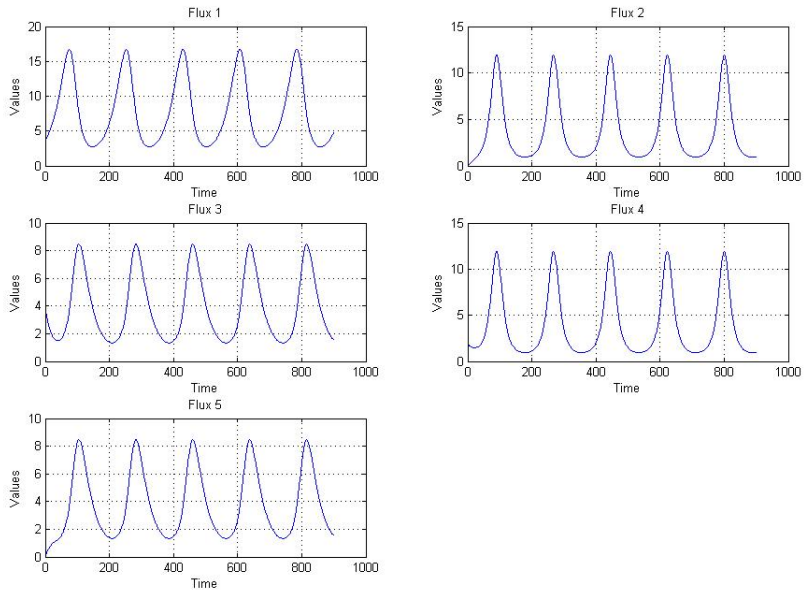

Fig. 1. The values of Sirius' fluxes calculated by using the flux regulation maps given in Table 2.

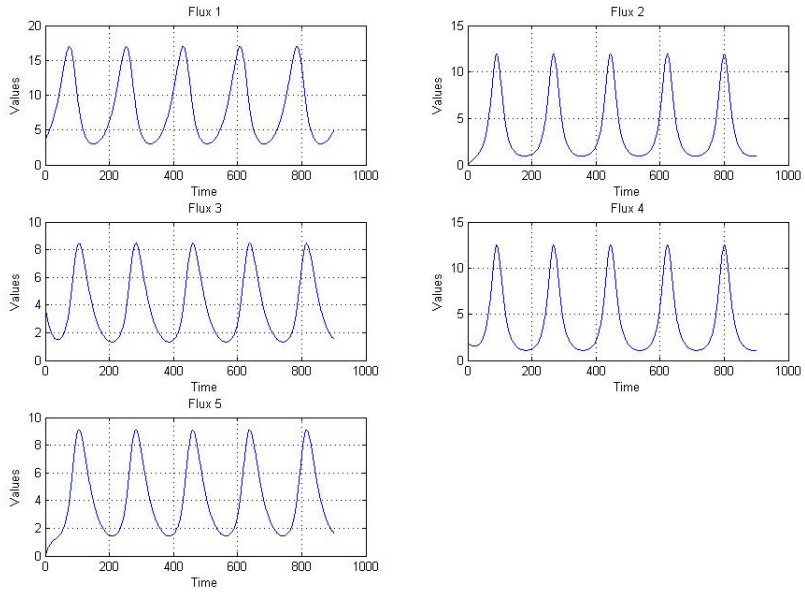

Fig. 2. The values of Sirius' fluxes calculated by applying the Log-gain theory and the initial vector of fluxes inferred by the proposed algorithm.

5.2 A biochemical case study

In this subsection the application of the algorithm to approximate the initial fluxes of the Belousov-Zhabotinsky reaction, also known as BZ reaction, is discussed. This reaction represents a famous example of a biochemical oscillatory phenomenon. Its importance is that it is the first evidence of a biochemical clock. Although the stoichiometry of the BZ reaction is quite complicated, several simplified mathematical models of this phenomenon have been proposed. In particular, Prigogine and Nicolis [15] proposed a simplified formulation of the dynamics of the BZ reaction, called *Brusselator*, whose oscillating behavior is represented by only two substances, x and y respectively, and it is governed by the following system of differential equations:

$$\begin{aligned} \frac{dx}{dt} &= k_1 - k_2 x + k_3 x^2 y - k_4 x \\ \frac{dy}{dt} &= k_2 x - k_3 x^2 y \end{aligned} \tag{14}$$

where $k_1 = 100, k_2 = 3, k_3 = 10^{-4}$ and $k_4 = 1$ represent constant rates. The numerical solution of the system system (14) from initial conditions $x = 1$ and

$y = 10$ shows the oscillatory dynamics displayed in Figure 3. We use this dynamics as experimental data on which applying our algorithm. By reading the set of

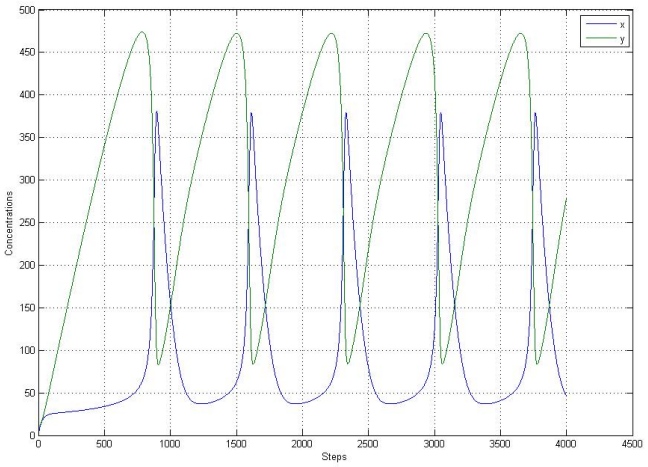

Fig. 3. Numerical solution of the system of differential equations (14).

differential equations (14) the stoichiometry of the Brusselator can be interpreted by using the set of rewriting rules reported in Table 3. In fact, species x has two positive and two negative contributions, while one positive and one negative contributions characterize y. Thus, the equations can be translate in the suitable stoichiometry by following the strategy described in [5].

Rules
$r_1 : \lambda \rightarrow x$
$r_2 : xxy \rightarrow xxx$
$r_3 : x \rightarrow y$
$r_4 : x \rightarrow \lambda$

Table 3. A set of rewriting rules that describes the Brusselator' stoichiometry.

In the case of BZ we adopt a different strategy of validation of our algorithm. In fact, there is a complete correspondence between the dynamics computed by the differential model and that one computed by the equational metabolic algorithm using the fluxes inferred (Figure 4) by solving an OLGA module starting from the initial fluxes inferred by means of our algorithm.

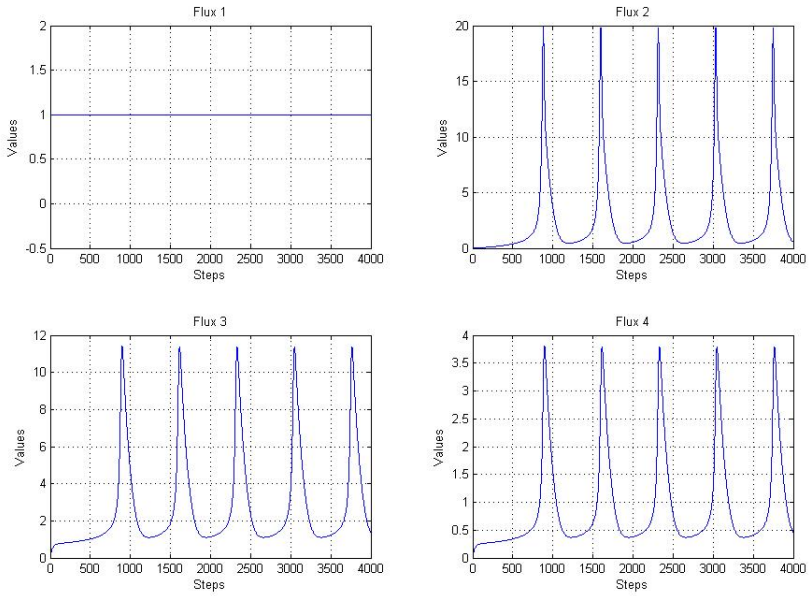

Fig. 4. The BZ reaction's fluxes calculated by using the Log-gain theory and the initial vector of fluxes inferred by our algorithm.

6 Conclusions

In this paper we have devised an algorithm for inferring the initial reaction fluxes of a metabolic network.

The proposed algorithm has been validated on test cases of a synthetic metabolic oscillator and Brusselator reaction. The near future investigations will be planed with the aim *i)* to show the applicability of our method to complex biological cases *ii)* and to improve this algorithm possibly with other relevant computational features.

References

1. P.B. Belousov: Sb. Ref. Radiats. Med. Medgiz. page 145, 1959.
2. L. von Bertalanffy: *General Systems Theory: Foundations, Developments, Applications*. George Braziller Inc, New York, NY, 1967.
3. G. Ciobanu, M.J. Pérez-Jiménez, G. Păun: *Applications of Membrane Computing (Natural Computing Series)*. Springer-Verlag, Berlin, 2006.
4. F. Fontana, L. Bianco, V. Manca: P systems and the modeling of biochemical oscillations. In R. Freud, G. Păun, G. Rozenberg, and A. Salomaa, editors, *Membrane Computing, WMC 2005,* LNCS 3850, Springer, 2005, 199–208.

5. F. Fontana, V. Manca: Discrete solution to differential equations by metabolic P systems. *Theoretical Computer Science*, 372 (2007), 165–182.
6. J.S. Huxley: *Problems of Relative Growth*. 2nd, Dover, New York, 1972.
7. D.S. Jones, B.D. Sleeman: *Differential Equations and Mathematical Biology*. Chapman & Hall/CRC, February 2003.
8. V. Manca: Log-Gain Principles for Metabolic P Systems. *Natural Computing*. To appear.
9. V. Manca: The Metabolic Algorithm for P systems: Principles and Applications. *Theoretical Computer Science*, 404 (2008), 142–157.
10. V. Manca: Fundamentals of metabolic P systems. In G. Păun, G. Rozenberg, and A. Salomaa, editors, *Handbook of Membrane Computing*, chapter 16. Oxford University Press, 2009. To appear.
11. V. Manca: Metabolic P dynamics. In G. Păun, G. Rozenberg, and A. Salomaa, editors, *Handbook of Membrane Computing*, chapter 17. Oxford University Press, 2009. To appear.
12. V. Manca, L. Bianco: Biological networks in metabolic P systems. *BioSystems*, 91, 3 (2008), 489–498.
13. V. Manca, L. Bianco, F. Fontana: Evolution and oscillation in P systems: Applications to biological phenomena. In G. Mauri, G. Păun, M. J. Pérez-Jiménez, G. Rozenberg, and A. Salomaa, editors, *Membrane Computing, WMC5*, LNCS 3365, Springer, 2005, 63–84.
14. V. Manca, R. Pagliarini, S. Zorzan: A photosynthetic process modelled by a metabolic P system. *Natural Computing*, 2008. DOI 10.1007/s11047-008-9104-x.
15. G. Nicolis, I. Prigogine: *Exploring Complexity. An Introduction*. Freeman ans Company, San Francisco, CA, 1989.
16. G. Păun: Computing with membranes. An introduction. *Bulletin of the EATCS*, 67 (February 1999), 139–152.
17. G. Păun: *Membrane Computing: An Introduction*. Springer, Berlin, 2002.
18. G. Păun, G. Rozenberg. A guide to membrane computing. *Theoretical Computer Science*, 287, 1 (2002), 73–100.
19. K.S. Scott: *Chemical Chaos*. Cambridge University Press, Cambridge, UK, 1991.
20. E.O. Voit: *Computational Analysis of Biochemical Systems*. Cambridge University Press, 2000.
21. P. Waage, C.M. Guldberg: Forhandlinger, *Videnskabs-selskabet i Christiana*, 35 (1864).
22. A.M. Zhabotinsky: Proc. Acc. Sci, USRR. 157 (1964), 392.